Building a domain-specific bilingual lexicon resource with Sketch Engine and Lexonomy: Taking Ownership of the Issues

Zaida Bartolomé-Díaz, Francesca Frontini

Universidad de Las Palmas de Gran Canaria
Calle Juan de Quesada, 30, 35001 Las Palmas de Gran Canaria, Las Palmas, Spain

Université Paul Valéry – Laboratoire Praxiling
Route de Mende, 34090 Montpellier - France

zaida.bartolome101@alu.ulpgc.es, francescafrontini@gmail.com

Abstract

Thanks to new technologies, the elaboration of specialized bilingual dictionaries can be made faster and more standardized, offering not only a dictionary of equivalents, but also the representation of a conceptual field.

Nevertheless, in view of these new tools and services, some of which are offered free of charge by European institutions, it is necessary to question the viability of their use by a lambda user and the previous knowledge required for such use, as well as the possible problems they may encounter.

In our communication we show a series of possible difficulties, as well as a methodological proposal and some solutions, by presenting an extract of a French-Spanish bilingual dictionary for the domain of architecture. The extract in question is a sample of about 30 terms created with the Lexonomy dictionary editor (Měchura 2017).

Keywords: bilingual lexicon, specialized lexicon, Lexonomy, architecture

1. Introduction

Due to the increasing process of globalization, professionals and specialized users need to communicate and use an appropriate terminology in each interlinguistic professional situation. Until now, the need to contrast terminological units adapted to each language pair and professional language required the simultaneous consultation of monolingual and bilingual dictionaries. Nowadays, as a result of the application of computational linguistics and corpus linguistics to language studies, it is undeniable that we are facing a new era of lexicography (Fuertes-Olivera 2012). Despite this, support for the creation of bilingual lexicographic resources dedicated to technical language is still insufficient.

With the emergence of online and born-digital dictionaries and resources and the new technologies that have emerged around the Lexicographic Linked Open Data cloud (LLOD), the horizon has broadened considerably, recently also due to initiatives such as the ELEXIS[1] project, which have been developed to provide lexicographers with services and tools to facilitate access to such services.

But it is essential to test specific cases and to verify the different existing technological tools in relation to the needs and knowledge of a lambda-user.

The objective of our paper is to present a practical case for the creation of a bilingual French-Spanish lexicon, specialized in the field of architecture, using the Lexonomy dictionary editor (Měchura 2017). Our goal is not only to create a dictionary of equivalents, but also to produce a resource that represents the linguistic and conceptual knowledge of our field and that we can, later, visualize on the Internet and connect to the Linked Data cloud, in order to promote the representation of lexicographic resources,

which are currently still under-represented (Declerck 2018).

For the realization of this resource we started from a specialized corpus extracted from the Web using Sketch Engine[2]. Sketch Engine is a tool for building and exploring corpus. Through its algorithms it analyzes the different texts and is able to identify and extract specific terminology. It is also designed for text analysis or text mining applications.

After a domain-specific terminology extraction, we used the One-click Dictionary function of Sketch Engine to create a first draft of the dictionary, which we then customize and adapt to our needs and to the requirements of current language standards (a TEI serialization of LMF[3]) in the Lexonomy interface.

We intend to show the main problems encountered when a lexicographer or translator, not necessarily knowledgeable of programming or coding languages, manually refines and curates such a resource, and the solutions we have found to solve these problems, such as working off-line the XML file or sharing different sub-entries for avoiding repeat information. (Měchura 2018).

We will try to suggest a methodology applicable to any other technical domain for which there are no available resources or for under-represented language pairs.

2. Field of study

Studies and resources dedicated to the lexicographical treatment of architectural and technical terms in general language dictionaries are relatively scarce.

This field of architecture is perhaps one of the least analyzed specialized languages in the framework of the LSP, despite the importance of the current architectural

---

[1] https://elex.is/

[2] https://www.sketchengine.eu/

[3] Following the recently proposed ISO standard: https://www.iso.org/standard/75411.html.

historical heritage in both France and Spain and the variety of the subsets that make up its discourse.

Thus, through a quantitative and qualitative analysis of the existing dictionaries dedicated to architecture, whether on paper, electronic or born-digital, , we have remarked the evident lack of lexicographical material dedicated to this field of study, and even less when it comes to bilingual dictionaries combining French and Spanish.

Indeed, for this language pair, we have found very few dictionaries of this type, only four French-Spanish bilingual dictionaries specialized in architecture, published in 1967, 1968, 1969 and 2001, and only on paper.

This scarcity of studies in the field, as well as of lexical and terminological works, often poses serious problems for specialized users of a technical sector who are also often not sufficiently familiar with dictionaries, their usual contents and basic rules of management.

Once this lexicographic gap was identified, we thought it necessary to develop a resource dedicated to this field of study that would be easy to use and that would facilitate the development and transfer of knowledge, technologies, solutions and pilot projects to be carried out among French and Spanish speaking professionals in the field of architecture.

Moreover, from a lexicographical point of view, organized storage in a lexical database allows for better management and maintenance of information, and facilitates the detection and correction of errors. (Fernández-Pampillón Cesteros and Matesanz del Barrio 2006).

3.   Methodology

3.1   Selection of Corpus

Given that the construction of specialized dictionaries is generally carried out with corpus-based methodologies, the first difficulty we encounter is the lack of specialized corpora in our field of study. In our case we have not found a specific corpus in the field of architecture, neither monolingual, nor bilingual (neither parallel nor comparable.)

For this reason, we have decided to create our own corpus of study, specialized in architecture, starting from Web data.

In this sense, the Web today represents a huge corpus within the reach of linguists interested in specific studies, whose needs are not met by traditional and/or existing corpora.

The term "Web as corpus" was first introduced in 2001 by Kilgarriff and, two years later, developed by the article of the same title by Kilgarriff and Grefenstette (2003) in which arguments in favor of using the Web as a corpus were presented.

The Web is a good place for the discovery of neologisms (Hundt, Nesselhauf, and Biewer 2007), language use from a non-normative register (Lüdeling, Evert, and Baroni 2007), or obsolete language use not found in a traditional or existing corpus (Renouf and Kehoe 2006).

In order to conform our own comparable corpus, we have decided to use the Sketch Engine tool.

Sketch Engine has the option to create a parallel, bilingual or multilingual corpus for a specific field.

However, in our case our goal was to extract information about the specialized terminology for each of the two languages and not simply the terms or texts translated from each of them into the other.

So, we decided from the beginning to eliminate the idea of working with a parallel bilingual corpus and decided to form two specialized corpora separately in order to extract the specific terminology from each one of them.

Using the corpus creation tools offered by Sketch Engine, we have built our corpora from a specific search on the Web:

- For French we used a search query based on terms contained in *FrenceTerme* [4], related to the domaine: HABITAT ET CONSTRUCTION / Architecture (administrateur, -trice de bâti immobilier modélisé, agence d'architecture, bâti-immobilier-modélisé, bureau d'études techniques, calepinage, coupe, création architecturale, stylisme architectural, détail d'exécution, élévation, section, tirage).
  Sketch Engine extracted a series of web pages and documents from the search. We then proceeded to analyze and clean up the different documents (removal of messages from forums, blogs, advertisements...). We finally formed a French corpus specialized in architecture, made up of 82 documents.

- In order to constitute the Spanish corpus, since an equivalent of FrenchTerme was unavailable, we used another function of Sketch Engine, and crawled the content of an online portal specialized in Spanish architecture (www.plataformaarquitectura.cl).
  From this platform, which is the most read platform in Spanish dedicated to architecture according to the numbers of visits in 2019, a corpus composed of 65 documents was created.

3.2   Terminological extraction

Once our two specialized corpora were formed, we performed the standard procedure for terminology extraction proposed by Sketch Engine, comparing each of the corpuses with a general corpus in Spanish and French; as a result we obtained a series of representative architectural terms for each of the languages.

Since the objective of our work is to obtain a sample of a lexicon as well as the methodology for its creation, we will concentrate here only on the first 15 words of each language.

3.3   Draft dictionary

Since we worked with two separate comparable corpora, we could not directly form a single dictionary that would include both languages.

So, using the One-Click dictionary function of Sketch Engine, we first created a first draft dictionary for French,

---

[4] http://www.culture.fr/franceterme

from which, as indicated in the previous paragraph, we retain only the first 15 words of our terminology extraction.

Then, we manually added to this dictionary the words extracted from our Spanish corpus.

In this way we obtained a first small bilingual dictionary of 30 entries, unrelated to each other.

One of the problems that we found in this step, apart from the difficulty of adding one by one each of the 15 terms from the Spanish corpus, was the impossibility of relate our dictionary entries back to the both of the corpora. Indeed, Lexonomy offers the possibility to link the created dictionary with its original corpus available in Skechtengine. However, in our case, since we work from two comparable corpora, we can only link one of them to the bilingual dictionary.

3.4 Download the draft dictionary

At this point and after correctly understanding the operation of Lexonomy, we thought it is necessary to download the xml file and continue working on it using an xml editor (Oxygen).

From our point of view Lexonomy can be simple and easy to use to write some small entries, but if you really want to use the application for a more substantial work, with a greater number of terms and an in depth editing of the structure of each entry, working with off-line source files becomes more efficient.

At the same time the drawback is that it limits the options for several users to work on the same file online, which could be useful in the academic domain to carry out collaborative work with students, as well as being an impediment to the realization of more professional content.

3.5 Formatting the downloaded xml file off-line

The first problem we encounter when downloading the xml file from Lexonomy and open it in Oxygen is that each entry is organized independently.

The root element is minimal and does not contain information about the lexicon.

```
<?xml version="1.0"?>
<z8tq8vrq2>
<entry lxnm:entryID="1"
xmlns:lxnm="http://www.lexonomy.eu/">
    <headword
    xml:space="preserve">sujétion</headword>
    <partOfSpeech
    xml:space="preserve">an</partOfSpeech>
    <sense>
     <translation xml:space="preserve">sujeción
     </translation>
     <translation xml:space="preserve">fixación
     </translation>
     <example>
        <source><h>Sujétions</h><h>relatives</h>à
        l'ensemble des terrassements</source>
        <target> XXX </target>
     </example>
```

```
    </sense>
    <sense>
     <translation>limitación</translation>
     <example>
        <source>L'Entrepreneur fait son affaire auprès
        des services compétents de toutes démarches,
        autorisations ou
        <h>autres</h><h>sujétions</h>ayant trait au
        chantier, il fait établir notamment les
        branchements et canalisations pour la distribution
        de l'eau, de l'électricité et du téléphone
        correspondant aux besoins de chantier. </source>
        <target>XXX</target>
     </example>
    </sense>
</entry>

<entry lxnm:entryID="2"
xmlns:lxnm="http://www.lexonomy.eu/">
    <headword
    xml:space="preserve">coffrage</headword>
    <partOfSpeech
    xml:space="preserve">n</partOfSpeech>
    <sense>
     <translation xml:space="preserve">encofrado
     </translation>
     <example>
        <source xml:space="preserve">
        La partie haute du bâtiment également réalisée en
        coffrage inclinés est aussi d'une grande
        complexité, là encore des...</source>
       <target xml:space="preserve">XXX</target>
     </example>
    </sense>
</entry>
```

Figure 1. Extract of the XML file downloaded from Lexonomy

While the access via Lexonomy facilitates the task of creating an online version of the dictionary accessible to everyone, having an off line version of the dictionary with standard xml modeling is an important desiderata for the project, as it allows this resource to be used for more complex tasks, such as NLP and linking to other resources.

Since our intention is that eventually publish the final resource also following the aforementioned ISO LMF standard, an attempt was made to restructure our file using Oxygen and then try uploading it into Lexonomy again to verify compatibility.

In the previous Language resource management — Lexical markup framework (LMF), approved on 2008-03-21[5] we can find the Machine Readable Dictionary (MRD) extension which provides a metamodel package for representing data stored in machine readable dictionaries. This extension supports electronic machine-readable dictionary access for both human use and machine processing.

Since the MRD extension is based on the LMF core package and the morphological extension, it is designed to

---

[5] https://www.iso.org/standard/37327.html

interchange data with other LMF extensions where applicable.

Due to copyright reasons it is not possible to show here the MRD class Diagram here. It can be consulted by the corresponding ISO standard.

As indicated in this norm in a bilingual MRD, the Equivalent class represents the translation equivalent of the word form managed by the Lemma class. The Equivalent class has in a zero to many aggregate associations with the Sense class, which allows the lexicon developer to omit the Equivalent class from a monolingual dictionary.

While modeling our dictionary is possible, the standard is currently under revision as a multi part standard, with a part being dedicated to an xml serialization which uses the TEI dictionary model as a basis[6]. At the time of writing this module is being finalized, but not many examples exist of bilingual dictionaries in the new format, the proposed modeling we outline here was defined by consulting with an expert[7] , and is illustrated by the example in Figure 2:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI                 xmlns="http://www.tei-c.org/ns/1.0"
type="lexicon">
  <teiHeader>
    <fileDesc>
      <titleStmt>
       <title>Lexique Bilingue Architecture</title>
      </titleStmt>
      <publicationStmt>
       <publisher>Zaida Bartolomé Diaz</publisher>
       <pubPlace>Montpellier</pubPlace>
       <date>2020</date>
       <availability><p>This is an open access work
licensed under a Creative Commons Attribution 4.0
International license.</p></availability>
       <ptr
target="http://digitalcommons.unl.edu/zeabook/55
"/>
      </publicationStmt>
      <sourceDesc><p>Born digital: no previous source
exists.</p></sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <entry xml:lang="fr" xml:id="LA_1"
      xmlns:lxnm="http://www.lexonomy.eu/">
        <form type="lemma">
         <orth>sujétion</orth>
         <gramGrp>
           <pos>n.</pos>
         </gramGrp>
        </form>
        <sense n="1">
        <cit type="example">
           <quote><emph>Sujétions</emph>
```

```xml
         <emph>relatives</emph>à l'ensemble des
         terrassements
         </quote>
      </cit>
      <xr xml:lang="es" type="translation">
          <quote>sujeción</quote>
          <quote>fixación</quote></xr>
      </sense>
      <sense n="2">
      <cit type="example">
          <quote>L'Entrepreneur fait son affaire auprès
          des services compétents de toutes démarches,
          autorisations ou
          <emph>autres</emph><emph>sujétions</em
          ph>ayant trait au chantier, il fait établir
          notamment les branchements et canalisations
          pour la distribution de l'eau, de l'électricité et
          du téléphone correspondant aux besoins de
          chantier.</quote>
      </cit>
      <xr type="translationEquivalent">
      <ref xml:lang="es">limitación</ref></xr>
      </sense>
  </entry>
```

Figure 2. First version of TEI-LMF format of download xml file

As it can be observed this representation of the entry "sujetion" is included in a general <TEI> node of type lexicon; the <header> contains metadata information while lexical entries are included under <text><body>.

For each of the entries the section "Context" is represented by <cit type=example"><quote>…, in which we extract of the source corpus can be listed and linked by Sketch Engine. Each entry has one or more senses, and these in turn may have a <translationEquivalent>.

In view of this representation we have decided to make different modifications with which we believe we are advancing in our project to improve and provide a more professional and useful product.

First, to go further we wanted to add the **definition** section. To do this, rather than provide a definition of its own we wanted to collect a series of links to different dictionaries so that the user can go directly to a definition of the term in question provided by a monolingual technical dictionary.

Secondly, in addition, if we wish to create a dictionary of equivalents with cross entries between different entries, the term included in the <xr type: "translationEquivalent"> section should also be a separate entry, if it is not included in the extracted terms corresponding to the other language.

In that sense, our dictionary will also be enriched by new entries that will arise as a result of the translation of each term into the other language.

The following figure shows the resulting xml file after our modifications:

---

[6] See some examples in: https://github.com/DARIAH-ERIC/lexicalresources/tree/master/Schemas/LMFinTEI%20Specification

[7] Laurent Romary, currently co-leader of the new LMF Part 4 and expert of TEI.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI            xmlns="http://www.tei-c.org/ns/1.0"
type="lexicon">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Lexique Bilingue Architecture</title>
      </titleStmt>
      <publicationStmt>
        <publisher>Zaida Bartolomé Diaz</publisher>
        <pubPlace>Montpellier</pubPlace>
        <date>2020</date>
        <availability><p>This is an open access work
licensed under a Creative Commons Attribution
4.0 International license.</p></availability>
        <ptr
target="http://digitalcommons.unl.edu/zeabook/5
5"/>
      </publicationStmt>
    <sourceDesc><p>Born digital: no previous source
exists.</p>
    </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <entry xml:lang="fr" xml:id="LA_1"
xmlns:lxnm="http://www.lexonomy.eu/">
        <form type="lemma">
          <orth>sujétion</orth>
          <gramGrp>
            <pos>n.</pos>
          </gramGrp>
        </form>
        <sense n="1">
          <xr xml:lang="es"
type="translationEquivalent">condición</xr>
          <def xml:lang="fr" type="CNRTL">
            <quote
type="https://www.cnrtl.fr/definition/suj%C3
%A9tion">Contrainte liée aux choses elles-
mêmes.</quote>
          </def>
          <cit type="example">
            -<quote><emph>Sujétions</emph>
            <emph>relatives</emph>à l'ensemble des
terrassements
            </quote>
          </cit>
        </sense>
        <sense n="2">
          <xr xml:lang="es"
type="translationEquivalent">limitación</xr>
          <def xml:lang="fr" type="Eyrolles">
            <quote type="https://www.editions-
eyrolles.com/Dico-
BTP/definition.html?id=8566"> Servitude
étant due. Dans un marché, les sujétions
représentent les travaux ou services à
effectuer.</quote>
          </def>
```

```xml
          <cit type="example">
            <quote>L'Entrepreneur fait son affaire auprès
des services compétents de toutes démarches,
autorisations ou
            <emph>autres</emph><emph>sujétions</emp
h>ayant trait au chantier, il fait établir
notamment les branchements et canalisations
pour la distribution de l'eau, de l'électricité et
du téléphone correspondant aux besoins de
chantier.</quote>
          </cit>
        </sense>
      </entry>
```

Figure 3. Entry "sujétion" re-defined

3.6   Importing this xml file to Lexonomy

As a next step we tried to import the file we have worked
off-line into Lexonomy again and in order to test for
compatibility.

When we tried to import the file back into Lexonomy the
first problem was that the entries were not independent of
each other. In order to work again on Lexonomy, each entry
must be listed independently.

In view of this, we should ask ourselves how the metadata
of a file should be defined in Lexonomy if we want this
dictionary to be indexed and linked to the external
resources and to the lexicographic cloud which projects
such as ELEXIS are trying to build.

At this point and due to multiple problems, that prevent
Lexonomy from a correctly recognizing of our file we have
decided to recreate directly our scheme in our profile.
Lexonomy doesn't recognize our file: it doesn't order the
entries correctly and the structure is lost when we upload
our document.

3.7   Recreating the TEI-LMF serialization directly in Lexonomy

As a result of our previous attempts, we decided to return
to our first dictionary, which included all 15 French terms
and all 15 Spanish terms.

The objective was to recreate the TEI-LMF serialization
that we had previously developed off-line directly within in
the Lexonomy interface. This of course required also an ad
hoc a style sheet.

However, recreating this serialization in the Lexonomy
interface poses several problems since the result is not
exactly what we expected.

Nevertheless, a positive outcome if this operation is that
through Lexonomy's interface we were able to create cross-
reference between the different French and Spanish terms,
which also allowed us to add terms (translations) that we
had not initially foreseen.

We also decided to collect the collocates and thesauruses
and make them sub-entries as recommended by Měchura
(2018), so that the information could also be reused.

In this respect, it should be noted that Lexonomy requests
that entries be re-indexed each time new sub-entries are
created. Such indexing is very time consuming and often

requires a restart of the application for the changes to be taken into account.

Figure 4. Recreating TEI-LMF serialization of an entry directly in Lexonomy

Figure 5. Style sheet of an entry defined directly in Lexonomy

4. Conclusions

After this first experience creating a dictionary of equivalents in a certain specialized field and for a couple of languages for which there are not yet enough resources, we can present several reservations:

- Regarding Sketch Engine we regret that there is no function or procedure to create a single, affordable corpus composed of different texts in different languages.

The fact that Sketch Engine only allows the creation of parallel corpus restricts its use, being limited to only possible use it when bilingual or multilingual parallel texts, or translations of a given corpus are available.

- Regarding to Lexonomy the first obstacle was, as expressed above, not being able to directly create a dictionary on Lexonomy through the one-click dictionary function from two different corpora, not even one that we can link ourselves subsequently.

We think that it could be interesting to contemplate the idea of creating a single dictionary in Lexonomy from two different corpora created in Sketch Engine.

This will also allow to improve the links to collocates or thesaurus items from these different corpora for different languages. As we can see in the example of the entry "sujétion" we have recovered from our French corpus linked the collocations and thesaurus items. However, when it comes to developing an entry in Spanish it will not be possible to link our corpus since only one is allowed.

As described in our methodology it was also impossible for us to make directly on Lexonomy a dictionary that will contemplate the TEI-LMF serialization, mainly due to the fact that it is not possible to form a root that gathers all the metadata and all the entries from a same lexicon.

As we have observed when we have downloaded our file from Lexonomy and opened it in an XML editor these metadata are non-existent and if we try to do the opposite, the Lexonomy editor does not recognize a common root for all the entries since each one of them must be independent.

In view of this, we must ask ourselves how the metadata of a Lexonomy dictionary will be recorded.

We believe that if we don't contemplate this common root that collects metadata and all the entries of a dictionary it will be complicated, if not impossible, that a dictionary created and stored in Lexonomy can be tracked, found and shared with other users.

Regarding the display of our dictionary in Lexonomy we can point out mainly the fact that external references cannot be directly displayed with a hyperlink.

Until now (possibly in June when the new version will be implemented) Lexonomy does not allow to directly include hyperlinks to external pages. It is therefore necessary to collect and display the full link.

Finally, we regret the fact that the sub-entries do not work properly, and it is often necessary to restart the application in order to reindex them.

In spite of our objections, we believe that the idea of creating this database- interface in which to store and visualize lexicographical data is really interesting and since

it is relatively recent we believe that it needs to continue evolving until it becomes fully operational.

We believe that experiences like the one presented above mostly serve to have a real awareness of the existing problems and to be able to take into account the feedbacks of the users' experiences.

5. Bibliographical References

Declerck, T. 2018. 'Towards a Linked Lexical Data Cloud Based on Ontolex-Lemon.P'. In *Proceedings of the LREC 2018 Workshop "6th Workshop on Linked Data in Linguistic (LDL-2018)"*, 91. Miyazaki, Japan: In John P. McCrae, et al., editors.

Fernández-Pampillón Cesteros, A., and M. Matesanz del Barrio. 2006. 'Los Diccionarios Electrónicos: Hacia Un Nuevo Concepto de Diccionario'. *Estudios de Lingüística Del Español* 24 (Nuevos géneros discursivos: los textos electrónicos). http://elies.rediris.es/elies24/pampillon_cap1.htm.

Fuertes-Olivera, P. 2012. 'Lexicography and the Internet as a (Re-)Source'. *Lexicographica* 28. https://doi.org/10.1515/lexi.2012-0005.

Hundt, M., N. Nesselhauf, and C. Biewer. 2007. 'Corpus Linguistics and the Web'. In *Corpus Linguistics and the Web*, In M. Hundt, N. Nesselhauf, C. Biewer (Eds.), 1–5. Amsterdam – New York: Rodopi.

Kilgarriff, A., and G. Grefenstette. 2003. 'Introduction to the Special Issue on the Web as Corpus'. *Computational Linguistics* 29 (September): 333–47. https://doi.org/10.1162/089120103322711569.

Lüdeling, Anke, Stefan Evert, and Marco Baroni. 2007. 'Using Web Data for Linguistic Purposes'. In *Corpus Linguistics and the Web*, edited by Marianne Hundt, Nadja Nesselhauf, and Carolin Biewer. Brill | Rodopi. https://doi.org/10.1163/9789401203791_003.

Měchura, M. 2017. 'Introducing Lexonomy: An Open-Source Dictionary Writing and Publishing System'. In *Electronic Lexicography in the 21st Century: Proceedings of ELex 2017 Conference*, 662–79, 18 pp. Leiden: In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek, V. Baisa.

Měchura, M. 2018. 'Shareable Subentries in Lexonomy as a Solution to the Problem of Multiword Item Placement'. In *The XVIII EURALEX International Congress: Lexicography in Global Contexts*, 223–32, 10 pp. Ljubljana, Slovenia: In. Jaka Čibej, Vojko Gorjanc, Iztok Kosem and Simon Krek.

Renouf, A., and A. Kehoe. 2006. . '. Introduction: The Changing Face of Corpus Linguistics'. In *The Changing Face of Corpus Linguistics*, In A. Renouf & A. Kehoe (Eds.). Amsterdam – New YorK: Rodopi.